Semi-Annual Performance Report
August 1, 1991 — January 31, 1992

Transform-Based Wideband Array Processing

Douglas B. Williams, Principal Investigator
Rabinder N. Madan, Scientific Officer

Contract No.: N00014-91-J-4129
R&T No.: 4148136- - -01

Over the past six months, this contract has funded two projects in full and one project in part. The two fully funded projects focus on the application of random coefficient models to wideband high-resolution direction finding and transient signal detection and estimation. The partially funded project involves the analysis of nonlinear, possibly chaotic, dynamical systems. Progress in each of these areas is described below.

Random Coefficient Models: The application of random coefficient models to narrowband high resolution direction finding has been very successful and has already resulted in a conference paper to be presented at ICASSP92 (see the attached paper by Jost & Williams). It has been shown that the random coefficient model is much better suited to modeling sensor array data than the autoregressive model is. A simple method for estimating the parameters of the random coefficient model has been developed and applied to simulated data. Finally, a beamformer for the random coefficient model has been developed which has significantly better performance then earlier linear predictive beamformers. A journal paper describing these results is currently being written and should be submitted to the *IEEE Transactions on Signal Processing* by this Spring. There is also progress being made toward the final goal of applying these techniques to wideband direction finding.

Transient Signal Detection: The application of the wavelet transform to the detection of transient signals with an array of sensors is being examined. This approach has led to a directional multirate filter bank structure that decomposes the incoming signal into decaying exponentials. This filter bank is also capable of adapting towards an improved estimate of the structure of the transient signal which, consequently, also improves the detection performance. Publication of these results will proceed after simulations and comparisons to other transient signal detectors are complete.

Nonlinear System Identification: System identification algorithms that depend on gradient descent methods have been found to degrade significantly if the time-series or, equivalently, the system that produced the time-series is chaotic (see the attached paper by Drake & Williams). A careful analysis of these degradations has led to algorithms which are much less sensitive to the potentially chaotic nature of these nonlinear systems. Analysis of these systems and their time-series continues with the eventual goal being a very general system identification algorithm to be applied to the time-series produced by nonlinear discrete-time systems.

To be presented at ICASSP, San Francisco, California, March 1992.

A BEAMFORMER BASED UPON THE RANDOM COEFFICIENT MODEL *

Bruce Jost *Douglas B. Williams*
Georgia Institute of Technology, Atlanta, GA, 30332-0250 USA

ABSTRACT

This paper applies the random coefficient model to array processing, specifically in the design of a beamformer for direction finding. This model is similar to the autoregressive (AR) model, except the coefficients are allowed to change with time instead of remaining constant; thus allowing the beamformer to better model any additive noise or signal correlations in the observations. Through the use of a binary hypothesis test, it is shown that random coefficient models better fit typical array data than do AR models. A Kalman filter is presented that has the array observations as inputs and the parameters of the random coefficient model as outputs. A new beamformer based on the random coefficient model is derived that is similar to the constant coefficient linear predictive (CCLP) beamformer. The two beamformers are compared and it is shown that the random coefficient beamformer outperforms the CCLP beamformer.

1. INTRODUCTION

The linear model is probably the most popular model in use in engineering and science because it is simple and yet powerful. Many times, computational advantages due to the simplicity of the linear model far outweigh any performance gain achieved by more complicated models. The commonly used autoregressive (AR) model is one such linear model. In array processing the AR model leads to linear predictive (LP) beamforming [1] where the output of a selected sensor, say the m_0th, is estimated as a weighted linear combination of the other sensor outputs. Assuming narrowband signals and letting $X_m(k), k = 0, \ldots, N-1$, be snapshots of the Fourier transform of the mth sensor's output at the signals' common center frequency, the LP model for $X_{m_0}(k)$ is

$$X_{m_0}(k) = -\sum_{m \neq m_0} a_m X_m(k), \qquad (1)$$

where $\{a_m\}$ is a set of complex-valued weights to be found.

When compared to other beamformers with roughly equivalent computational complexity, such as the minimum variance distortionless receiver, the LP method possesses superior resolution properties [1]. However, the performance of the LP beamformer is highly dependent on both the signals' directions of propagation and the selection of m_0. Also, in many cases the direction-of-arrival estimation bias is quite high. The primary causes of this estimation bias are additive noise in the observations and correlations between propagating signals. Additive noise and signal correlations do not obey an AR relationship across the array; so the model in (1) is no longer accurate.

In order to keep the simplicity of the linear model of (1) and yet more accurately match the array data, the random coefficient model is proposed. This model is identical to the AR model except the weights a_m are random instead of constant and (1) becomes

$$X_{m_0}(k) = -\sum_{m \neq m_0} a_m(k) X_m(k). \qquad (2)$$

The coefficients can be expressed as

$$a_m(k) = \beta_m + \nu_m(k), \qquad (3)$$

where β_m is the mean of the coefficients for the mth sensor and the $\nu_m(k)$ are zero-mean independent identically distributed random variables. The random coefficient model of (2) keeps the simple linear form of the AR model and, hence, its computational advantages. Since the constant coefficients of an AR model can be expressed in terms of (3) with the $\nu_m(k)$ equal to zero, the random coefficient model can actually be thought of as a generalization of the AR model. Therefore, the random coefficient model will fit the array data at least as well as the AR model and will simplify to the AR model when it is the better choice.

2. TESTING OF THE MODELS

Through the use of a binary hypothesis test developed by Breusch and Pagan [2], it is possible to test which model, AR or random coefficient, will better fit typical array data. The test indicates that if the residuals from an AR model have a constant variance, an AR model is better suited; otherwise, a random coefficient model should be considered. The test is performed by generating a Lagrange multiplier (LM) statistic from generated array data. If the null hypothesis of an AR model is true, then the LM statistics will be asymptotically distributed as chi-squared with M degrees of freedom, where M is the number of sensors in the array. The method used for this paper was to generate

*THIS WORK WAS SUPPORTED BY THE OFFICE OF NAVAL RESEARCH UNDER CONTRACT NUMBER N00014-91-J-4129

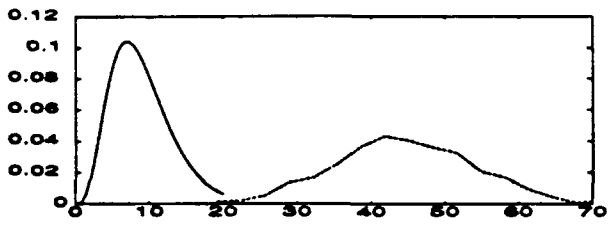

Figure 1: Results of the hypothesis test. The ten degree of freedom chi-squared distribution is the solid line while the normalized histogram of the LM statistics is the broken line. The array data was from a uniform linear array of 10 sensors with spacing $\frac{\lambda}{2}$. There are two signals present at $-10°$ and $30°$, the noise is additive white Gaussian and the SNR is 6 dB. There are 100 independent samples being used ($N = 100$).

array data for a given number of signals at their respective angles of arrival, and a given number of sensors, time samples, and signal-to-noise ratio. The LM statistic given in [2] was calculated for a large number of sets of data and the normalized histogram of these LM statistics was calculated and plotted against a chi-squared density function with M degrees of freedom to see how closely the two matched. Figure 1 shows an example of this test in which the normalized histogram of the LM statistics clearly does not match the chi-squared density plot. In fact, it was found that the AR hypothesis is rejected for even the simplest cases if any additive noise or signal correlation is present. Therefore, it would seem that the random coefficient model is better suited for array processing applications.

3. RANDOM COEFFICIENT BEAMFORMER

As the random coefficient model has been shown to fit the data more accurately than the AR model, a beamforming algorithm based on this model has been developed. The remainder of the paper will present the algorithm for estimating the random coefficients and then the derivation of the random coefficient beamformer.

3.1. Estimating the Coefficients

The first step is to estimate the random coefficients. This is performed with the Kalman filter outlined in [3] and [4]. The following set of vectors is introduced to ease the computations:

$$\begin{aligned}
\mathbf{a}'(k) &= [a_0(k)\cdots a_{m_0-1}(k)\, a_{m_0+1}(k)\cdots a_{M-1}(k)] \\
\mathbf{b}' &= [\beta_0\cdots\beta_{m_0-1}\,\beta_{m_0+1}\cdots\beta_{M-1}] \\
\mathbf{v}'(k) &= [\nu_0(k)\cdots\nu_{m_0-1}(k)\,\nu_{m_0+1}(k)\cdots\nu_{M-1}(k)] \\
\mathbf{z}'(k) &= [X_0(k)\cdots X_{m_0-1}(k)\,X_{m_0+1}(k)\cdots X_{M-1}(k)]
\end{aligned}$$

where M is the number of sensors and ($'$) denotes Hermitian transpose. An error covariance matrix is defined as

$$\Sigma = [\sigma_{ij}^2],\ i,j = 0,\ldots,m_0-1,m_0+1,\ldots,M-1$$

where $E\{\nu_i(k)\nu_j^*(l)\} = \delta(k-l)\sigma_{ij}^2$. From [3] and [4] the estimates of $\mathbf{a}(k)$ are calculated from the Kalman filter equations

$$\begin{aligned}
\hat{\mathbf{a}}(k/k-1) &= \mathbf{b} \\
\hat{\Sigma}(k/k-1) &= \Sigma \\
h(k) &= \mathbf{z}'(k)\hat{\Sigma}(k/k-1)\mathbf{z}(k) + \sigma_n^2 \\
\hat{\mathbf{a}}(k/k) &= \mathbf{b} + \hat{\Sigma}(k/k-1)\mathbf{z}(k)h^{-1}(k) \\
&\quad \cdot [X_{m_0} + \mathbf{z}'(k)\mathbf{a}(k/k-1)] \\
\hat{\Sigma}(k/k) &= \hat{\Sigma}(k/k-1) - \\
&\quad \hat{\Sigma}(k/k-1)\mathbf{z}(k)h^{-1}(k)\mathbf{z}'(k)\hat{\Sigma}(k/k-1) \quad (4)
\end{aligned}$$

where $\hat{\mathbf{a}}(k/k)$ is the estimate of $\mathbf{a}(k)$ at time k given all the observations up to k, $\hat{\Sigma}(k/k)$ is the error covariance matrix estimate of $\mathbf{a}(k)$ at time k given all the observations up to k, and σ_n^2 is the noise variance estimate.

To implement the Kalman filter, initial estimates of $\mathbf{b}, \Sigma$, and σ_n^2 must be calculated. The authors will present a method that calculates the least-squares estimates of these terms following [4]. Introducing a noise term into (2) yields

$$X_{m_0}(k) = -\sum_{m\neq m_0} a_m(k)X_m(k) + n_{m_0}(k) \qquad (5)$$

where n_{m_0} is white Gaussian noise and is uncorrelated with the observations $X_m(k)$. The noise associated with an arbitrary sensor, say the mth, has the property $E\{n_m(j)n_m^*(k)\} = \delta(j-k)\sigma_n^2$. Using (3), equation (5) can be written as

$$\begin{aligned}
X_{m_0}(k) &= -\sum_{m\neq m_0}\beta_m X_m(k) - \sum_{m\neq m_0}\nu_m(k)X_m(k) \\
&\qquad + n_{m_0}(k) \\
&= -\sum_{m\neq m_0}\beta_m X_m(k) + u(k) \qquad (6)
\end{aligned}$$

where

$$u(k) = n_{m_0}(k) - \sum_{m\neq m_0}\nu_m(k)X_m(k). \qquad (7)$$

Defining

$$\begin{aligned}
\mathbf{u}' &= [u(0)\cdots u(N-1)] \\
\mathbf{y}' &= [X_{m_0}(0)\cdots X_{m_0}(N-1)] \\
\mathbf{Z}' &= [\mathbf{z}(0)\cdots\mathbf{z}(N-1)],
\end{aligned}$$

equation (6) can be written as

$$\mathbf{y} = -\mathbf{Z}\mathbf{b} + \mathbf{u}.$$

Using ordinary least squares, the estimate of $\mathbf{b}$ is

$$\hat{\mathbf{b}} = -(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}. \qquad (8)$$

Taking the expected value of the magnitude squared of (7) yields

$$E\{|u(k)|^2\} = E\{|\, n_{m_0}(k) - \sum_{m\neq m_0}\nu_m(k)X_m(k)\,|^2\}$$

which reduces to

$$E\{|u(k)|^2\} = \sigma_n^2 + \sum_{i\neq m_0}\sum_{j\neq m_0}E\{\nu_i(k)\nu_j^*(k)X_i(k)X_j^*(k)\} \qquad (9)$$

where the other terms are zero because the noise is uncorrelated with the signals and coefficients. Assuming the $X_i(k)$ and $a_i(k)$ are zero mean and Gaussian, (9) can be expressed as

$$E\{|u(k)|^2\} = \sigma_n^2 + \sum_{i\neq m_0}\sum_{j\neq m_0}[E\{\nu_i(k)\nu_j^*(k)\}E\{X_i(k)X_j^*(k)\} + E\{\nu_i(k)X_i(k)\}E\{\nu_j^*(k)X_j^*(k)\} + E\{\nu_i(k)X_j^*(k)\}E\{\nu_j^*(k)X_i^*(k)\}]$$

which reduces to

$$E\{|\,u(k)\,|^2\} = \sigma_n^2 + \sum_{i\neq m_0}\sum_{j\neq m_0} E\{\nu_i(k)\nu_j^*(k)\}E\{X_i(k)X_j^*(k)\}. \tag{10}$$

Since

$$|\,u(k+l)\,|^2 = E\{|\,u(k+l)\,|^2\} + \xi_l \tag{11}$$

where the ξ_l is a zero mean error with unknown variance, (10) can be used in (11) as

$$|u(k+l)|^2 = \sigma_n^2 + \sum_{i\neq m_0}\sum_{j\neq m_0} E\{\nu_i(k+l)\nu_j^*(k+l)\}E\{X_i(k+l)X_j^*(k+l)\} + \xi_l$$

or in vector notation as

$$\mathbf{e} = \mathbf{X}\mathbf{s} + \mathbf{r}$$

where $\mathbf{e}' = [|u(0)|^2 \cdots |u(N-1)|^2], \mathbf{r}' = [\xi_0 \cdots \xi_{N-1}], \mathbf{s}' = [\sigma_n^2\ \sigma_{1,1}^2\ \sigma_{1,2}^2 \cdots \sigma_{1,M-1}^2\ \sigma_{2,1}^2 \cdots \sigma_{M-1,M-1}^2]$, and

$$\mathbf{X} = \begin{bmatrix} 1 & X_1(0)X_1^*(0) & X_1(0)X_2^*(0) & \cdots & X_1(0)X_{M-1}^*(0) & X_2(0)X_1^*(0) & \cdots & X_{M-1}(0)X_{M-1}^*(0) \\ \vdots & \vdots & & & & & & \vdots \\ 1 & X_1(N-1)X_1^*(N-1) & & \cdots & & & & X_{M-1}(N-1)X_{M-1}^*(N-1) \end{bmatrix}$$

Using ordinary least squares, the estimate of $\mathbf{s}$ is

$$\hat{\mathbf{s}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}, \tag{12}$$

from which the estimates of σ_n^2 and Σ are extracted.

3.2. The Beamformer

Once the random coefficients have been estimated, they can be used in a beamformer to calculate the direction(s) of the measured signal(s). Following the derivation of the constant coefficient LP beamformer, (5) can be rewritten as

$$\begin{aligned} n_{m_0}(k) &= X_{m_0}(k) + \sum_{m\neq m_0} a_m(k)X_m(k) \\ &= \sum_{m=0}^{M-1} a_m(k)X_m(k) \end{aligned} \tag{13}$$

if $a_{m_0}(k) = 1$ for $k = 0, \ldots, N-1$. Multiplying (13) with a similar expression for $n_{m_0+l}^*(k)$ and taking the expected value gives (dropping the time index for brevity)

$$E\{n_{m_0}n_{m_0+l}^*\} = \delta(l)\sigma_n^2 = \sum_i\sum_j E\{a_ia_j^*X_iX_{j+l}^*\}.$$

The right-hand expected value can be expanded as in (9)

$$\delta(l)\sigma_n^2 = \sum_i\sum_j E\{a_ia_j^*\}E\{X_iX_{j+l}^*\} + A(l) + B(l) \tag{14}$$

where

$$\begin{aligned} A(l) &= \sum_i\sum_j E\{a_iX_i\}E\{a_j^*X_{j+l}^*\} \\ B(l) &= \sum_i\sum_j E\{a_iX_{j+l}^*\}E\{a_j^*X_i\} \end{aligned}$$

for $l = -(M-1), \ldots, M-1$. The expected values in (14) can be estimated by averaging over the available data and coefficient estimates, e.g., $E\{a_j^*X_i\} = \frac{1}{N}\sum_{k=0}^{N-1}\hat{a}_j^*(k)X_i(k)$.

The spatial Fourier transform of (14) can be calculated for any array geometry, but if a linear uniform array is used, the transform can be parameterized in terms of just the direction of look θ. Therefore, the random coefficient beamformer for a uniform linear array steered in direction θ can be written as

$$P^{RC}(\theta) = \frac{\sigma_n^2 - \mathcal{A}(\theta) - \mathcal{B}(\theta)}{\sum_{i=1}^{M}\sum_{j=1}^{M}E\{a_ia_j^*\}e^{-j2\pi\frac{d}{\lambda}(i-j)\sin\theta}} \tag{15}$$

where d is the spacing between sensors, λ is the wavelength of the propagating signals, and $\mathcal{A}$ and $\mathcal{B}$ are the spatial Fourier transforms of $A(l)$ and $B(l)$, respectively, e.g.,

$$\mathcal{A}(\theta) = \sum_{l=-(M-1)}^{M-1} A(l)e^{-j2\pi\frac{d}{\lambda}l\sin\theta}.$$

3.3. Simulation Results

The random coefficient beamformer of (15) can be compared to a constant coefficient LP (CCLP) beamformer. A typical example of this comparison is shown in figure 2 where the predictive element m_0 is the first sensor for both beamformers. It is seen that while both have sharp peaks at $-10°$ and $30°$, the random coefficient beamformer has a lower noise floor with none of the spurious peaks that might be confused as signals. This improvement is due to the random coefficient model being able to fit the noisy array data better than the AR model. As was expected from the results of the hypothesis test, in the comparisons run by the authors the random coefficient beamformer clearly outperforms the CCLP almost every time and always performs at least as well as the CCLP.

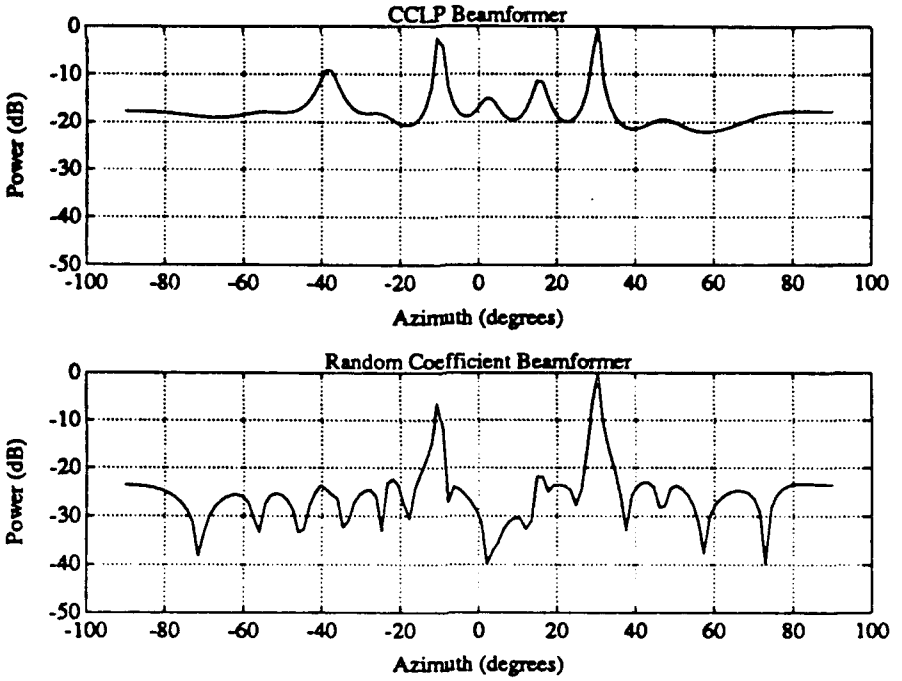

Figure 2: Comparison of the CCLP and random coefficient beamformers. The array data used was the same as that described in figure 1.

4. CONCLUSION

The random coefficient model was introduced for use in array processing. It was shown through the use of a hypothesis test that the random coefficient model will better fit typical array data than the constant coefficient autoregressive model. A beamformer using the random coefficient model was developed (15) as well as the method for estimating the random coefficients using a Kalman filter. The random coefficient beamformer was then compared to the familiar constant coefficient linear predictive beamformer and was shown to be a dramatic improvement.

REFERENCES

[1] D. H. Johnson, "The application of spectral estimation methods to bearing estimation problems," *Proc. IEEE*, vol. 70, no. 9, pp. 1018–1028, 1982.

[2] T. S. Breusch and A. R. Pagan, "A simple test for heteroscedastity and random coefficient variation," *Econometrica*, vol. 47, pp. 1287–1294, Sept. 1979.

[3] P. Newbold and T. Bos, *Stochastic Parameter Regression Models*. Beverly Hills, CA: Sage Publications, 1985.

[4] A. S. Abutaleb, "Adaptive line enhancement using a random AR model," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1211–1215, July 1990.

[5] D. F. Nicholls and B. G. Quinn, *Random Coefficient Autoregressive Models: An Introduction*. New York: Springer-Verlag, 1982.

[6] C. Hildreth and J. P. Houck, "Some estimators for a linear model with random coefficients," *J. of Amer. Stat. Ass.*, vol. 63, pp. 584–595, June 1968.

To be presented at ICASSP, San Francisco, California, March 1992.

ON ERROR FUNCTION SELECTION FOR THE ANALYSIS OF NONLINEAR TIME SERIES

Daniel F. Drake and Douglas B. Williams

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, GA 30332

ABSTRACT

The extreme sensitivity of a chaotic system's steady state response to small changes in its initial conditions makes long term prediction of the evolution of such a system difficult, if not impossible. In the framework of parameter estimation, we show how this sensitivity can hinder attempts to determine model parameters that will reproduce a target chaotic time sequence. Specifically, a waveform error minimization technique based on gradient descent optimization is not well suited for estimating the parameters of a strongly chaotic system. We propose a modification of this minimization procedure that avoids some of the obstacles present when estimating the parameters of a chaotic system.

1. INTRODUCTION

Chaos—unpredictable deterministic behavior—has been observed in phenomena ranging from chemical reactions [1] to solar flares [2]. Modelling time sequences derived from such processes can provide insight into the underlying physics that drive them. Unfortunately, the intrinsic sensitivity of chaotic systems makes them difficult to model; a representation with enough freedom to correctly reproduce chaotic behavior will itself be extremely susceptible to small variations in its parameters.

The realization that long-term prediction of certain completely deterministic systems was impossible sparked interest in a new area of Dynamical Systems, an area dealing with the phenomenon of chaos. The classic description of a chaotic system usually includes the phrase "sensitive dependence on initial conditions" [3]. Sensitive, in this context, refers to the exponential rate at which initially close trajectories on the attractor diverge. This sensitivity can be quantified by the spectrum of Lyapunov exponents associated with the attractor.

Quatieri and Hofstetter [4] wished to determine the parameters and initial conditions of a nonlinear difference equation whose solution would be as close as possible to some target time sequence generated by a dynamical system. They derived a gradient descent method that minimized the waveform error between the solution of the difference equation model and the target sequence.

We will show that the waveform error surface is not well-behaved if the target sequence is generated by a chaotic system. In particular, in the neighborhood of the global minimum at least one eigenvalue of the Hessian of the waveform error increases exponentially as a function of the length of the target sequence. This unbounded growth sets the global minimum at the bottom of a deep trench, rendering gradient descent techniques impractical. Consequently, we have modified the waveform error minimization procedure by taking into account the behavior of the error surface as a function of target sequence length. This improved optimization technique provides a much wider basin of attraction for the global minimum than the original method.

This work was supported in part by the Office of Naval Research under contract N00014-91-J-4129.

2. DYNAMICAL SYSTEMS AND CHAOS

Suppose $\mathbf{h} : \mathbf{M} \times \mathbf{R}^k \rightarrow \mathbf{M}$ is a parameterized discrete-time dynamical system defined on a smooth compact manifold $\mathbf{M}$ such that

$$\mathbf{y}[n] = \mathbf{h}(\mathbf{y}[n-1], \mathbf{p}). \tag{1}$$

We assume that this system is stable, and further that the state $\mathbf{y}[n]$ converges onto an attractor $\Lambda \subset \mathbf{M}$ as n tends to infinity for any initial condition $\mathbf{y}[-1]$ contained in the basin of attraction of Λ.

It can be shown that the variation of the state $\mathbf{y}[n]$ with respect to initial conditions $\mathbf{y}[-1]$ is given by

$$\mathbf{D}_{\mathbf{y}[-1]}\mathbf{y}[n] = \prod_{i=-1}^{n-1} \mathbf{D}_{\mathbf{y}}\mathbf{h}(\mathbf{y}[i], \mathbf{p}). \tag{2}$$

where the operator $\mathbf{D}_{\mathbf{x}}$ applied to the vector-valued function $\mathbf{f}(\mathbf{x})$ results in a matrix with elements $(\mathbf{D}_{\mathbf{x}}\mathbf{f})_{i,j} = \partial f_i(\mathbf{x})/\partial x_j$. The Lyapunov exponents quantify the average rate at which small perturbations of the initial condition are exponentially amplified or attenuated upon iterations of the system [5]. An infinitesimal deviation $d\mathbf{y}[-1]$ will result in a deviation

$$d\mathbf{y}[n] = \prod_{i=-1}^{n-1} \mathbf{D}_{\mathbf{y}}\mathbf{h}(\mathbf{y}[i-1], \mathbf{p}) d\mathbf{y}[-1], \tag{3}$$

and for almost all $d\mathbf{y}[-1]$

$$\|d\mathbf{y}[n]\| \approx e^{\lambda(n+1)} \|d\mathbf{y}[-1]\|, \tag{4}$$

where λ is the largest Lyapunov exponent of $\mathbf{h}$ on Λ. Eq (4) is equivalent to saying that $\prod_{i=-1}^{n-1} \mathbf{D}_{\mathbf{y}}\mathbf{h}(\mathbf{y}[i-1], \mathbf{p})$ has an

eigenvalue that grows on average and in absolute value as $e^{\lambda(n+1)}$. A system is, by definition, chaotic if it has least one positive Lyapunov exponent, indicating its exponential sensitivity to small variations in initial conditions.

Similarly one can show that the dependence of the state on small variations of the system's parameters is given by

$$\mathbf{D_p y}[n] = \sum_{i=-1}^{n-1} \left(\prod_{j=i+1}^{n-1} \mathbf{D_y h}(\mathbf{y}[j], \mathbf{p}) \right) \mathbf{D_p h}(\mathbf{y}[i], \mathbf{p}). \quad (5)$$

The term $\mathbf{D_p h}$ converts small deviations in the parameters into small deviations in the state which are then propagated forward by the product $\prod \mathbf{D_y h}$. This coupling between parameters and state implies that a chaotic system will be extremely sensitive not only to variations in its initial conditions but to variations in its parameters as well.

3. WAVEFORM ERROR MINIMIZATION

Suppose we have a scalar time sequence $x[n] = v(\mathbf{y}[n])$ derived from a dynamical system via $v : M \rightarrow \mathbf{R}$. We assume this sequence is the solution of an m^{th} order nonlinear difference equation with a known form, but depending on k unknown parameters $\mathbf{p}$. An estimate $\hat{\mathbf{p}}$ of these parameters produces the time sequence estimate

$$\hat{x}[n] = f(\hat{\mathbf{x}}[n-1], \hat{\mathbf{p}}) \text{ with } 0 \leq n < N, \quad (6)$$

where $\hat{\mathbf{x}}[n-1] = (\hat{x}[n-1], \hat{x}[n-2], \ldots, \hat{x}[n-m])^T$ is the vector of the last m values of x at time n. For simplicity we assume that the initial conditions $\mathbf{x}[-1]$ are known exactly[1]; let $\hat{\mathbf{x}}[-1] = \mathbf{x}[-1]$. We wish to find the parameters that minimize the waveform error

$$E_N = \frac{1}{N} \sum_{n=0}^{N-2} (\hat{x}[n] - x[n])^2. \quad (7)$$

Quatieri and Hofstetter use a gradient descent method to minimize the waveform error with respect to parameters. An initial estimate $\hat{\mathbf{p}}$ of the parameter values is iteratively updated

$$\hat{\mathbf{p}} \leftarrow \hat{\mathbf{p}} - \mu (\mathbf{D_p} E_N)^T \quad (8)$$

so that the error decreases at each step. The step size μ is chosen so that the error decreases at each iteration; the minimization procedure terminates when the waveform error falls below a specified threshold.

In order to better understand the behavior of the error surface, we expand E_N about the global minimum $\mathbf{p}$:

$$E_N \approx \frac{1}{N} d\mathbf{p}^T \left(\sum_{n=-1}^{N-2} (\mathbf{D_p} x[n])^T (\mathbf{D_p} x[n]) \right) d\mathbf{p}, \quad (9)$$

where $d\mathbf{p}$ represents an infinitesimal deviation from the true parameters. A remarkable result by Takens [6] states that under the proper conditions, a scalar time sequence can be 'time delay embedded' into $\mathbf{R}^m$, revealing a diffeomorphic copy of the phase space dynamics that generated the sequence. The embedding is represented by the sequence of vectors $(\mathbf{x}[n])$, where $\mathbf{x}[n] = (x[n], x[n-1], \ldots, x[n-m+1])$, with the embedding dimension m suitably chosen . The diffeomorphic relationship between the true trajectory in phase space and the reconstructed one preserves certain quantities, namely the Lyapunov exponents. Therefore the dynamical system

$$\mathbf{x}[n] = \mathbf{g}(\mathbf{x}[n-1], \mathbf{p}) = \begin{bmatrix} f(\mathbf{x}[n-1], \mathbf{p}) \\ x[n-1] \\ x[n-2] \\ \vdots \\ x[n-m+1] \end{bmatrix} \quad (10)$$

has the same Lyapunov exponents as the original dynamical system. The gradient of the scalar time sequence $(x[n])$ is simply the first row of the matrix $\mathbf{D_p x}[n]$. If the dynamical system that produced the sequence is chaotic, then $\|\mathbf{D_p} x[n]\|$ will grow exponentially fast with increasing n, since it generally won't be orthogonal to the eigenvector along which the exponential expansion is taking place. Thus the Hessian of the waveform error in Eq (9), composed of a sum of outer products of the vectors $(\mathbf{D_p} x[n])$, has an increasingly large eigenvalue. As N increases the global minimum will become sandwiched between two increasingly steep walls— not ideal conditions for gradient descent optimization.

As will be seen in the next section, the waveform error seems well-behaved for *short* chaotic sequences; the exponential amplification of parameter mismatch has little time *over* which to markedly modify the sequence estimate. Our modification of the waveform error minimization procedure takes advantage of this phenomenon. Instead of trying to optimize our parameter estimates for the whole target sequence at once, we sequentially minimize the waveform errors $E_0, E_1, \ldots, E_N$. Once E_n sinks below some fixed threshold, we repeat the minimization process on E_{n+1}, using the last estimates of the parameters as the initial guess for the next step. Since each error surface has the global minimum in common, successive minimization of the errors forces the parameter estimates closer to their true value. We've effectively expanded the global minimum's basin of attraction to that of a length-one sequence, independent of the true sequence length.

4. EXAMPLES

A simple system capable of exhibiting chaotic behavior is the logistic equation

$$x[n] = p_1 x[n-1](1 - x[n-1]), \quad (11)$$

which is both a scalar dynamical system and a first order nonlinear difference equation. The model exhibits markedly different types of steady state behavior depending on the choice of p_1. A parameter value of $p_1 = 3.5$ results in a period-four oscillation; in contrast, a value $p_1 = 3.7$ produces chaotic behavior. Figure 1 contrasts the waveform error for both cases, for sequences of length $N = 87$ and an initial condition $x[-1] = 0.42$. While relatively flat and smooth in the periodic case, the waveform error in the chaotic case is riddled with local minima and the global

[1] The origin of the sequence can always be shifted to the right by m samples.

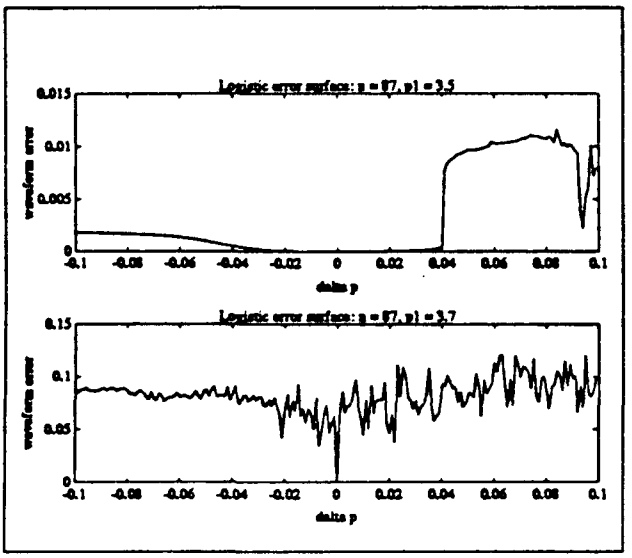

Figure 1: The waveform errors of periodic (above) and chaotic (below) target sequences of length 87, generated by the logistic equation.

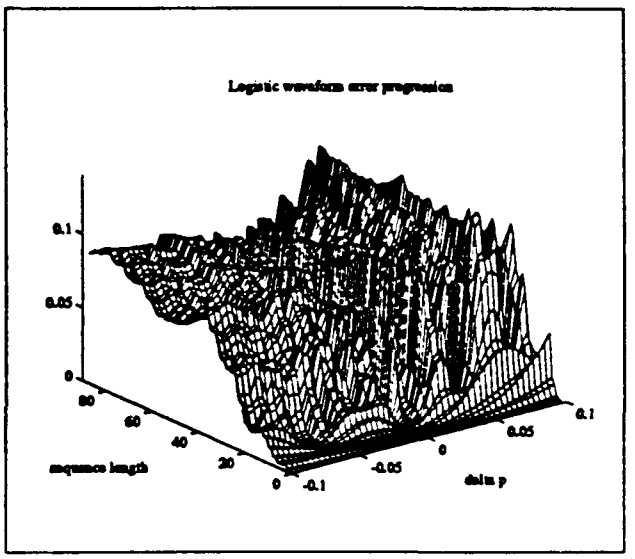

Figure 2: The waveform error as a function of target sequence length. In this case the sequence was generated by the chaotic logistic equation. The error surface is relatively smooth and shallow for short sequences, and becomes increasingly rough as more of the sequence is considered.

minimum has the very narrow basin of attraction as discussed above.

Figure 2 illustrates the behavior of the waveform error for the chaotic logistic system with respect to the sequence length. As noted previously, error surfaces for short sequence lengths are relatively smooth, giving parameter estimates with relatively large errors a greater chance of converging to the true parameter.

Figure 3 compares the performance of the original waveform error minimization technique, which tries to minimize the waveform error E_N directly, and our modified version. As expected from the general appearance of the waveform error, an initial deviation of only 5×10^{-8} in the model parameter gets trapped almost immediately in a local minimum, and the error never descends below the specified threshold of 10^{-5}. Our extension method, on the other hand, correctly identifies the true parameter after extending the target sequence to $N = 87$, even though the initial deviation from the true parameter value was 5×10^{-2}; six orders of magnitudes larger. In fact, any initial parameter value within the logistic equation's usual working range $p_0 \in [0, 4]$ will be converge to the the true parameter value.

An example of a two-parameter chaotic system is the Hénon system

$$y_1[n] = 1 - p_1 y_1^2[n-1] + y_2[n-1] \quad (12)$$

$$y_2[n] = p_2 y_1[n-1] \quad (13)$$

with parameter values $p_1 = 1.4$, $p_2 = 0.3$. If we consider the time sequence produced by the first variable ($y_1[n]$) our difference equation model has the form

$$x[n] = 1 - p_0 x^2[n-1] + p_1 x[n-2], \quad (14)$$

with initial conditions $x[-1] = y_1[-1] = 0.948586$ and $x[-2] = y_2[-1]/p_2 = 0.425317$.

Figure 4 shows that our extension method outperforms the original waveform error minimization technique. Initial parameter estimates with errors of more that 10^{-2} are reduced by six orders of magnitude. The original method, initiated with deviations of only 5×10^{-8} in both parameters, is immediately trapped in a local minimum.

Unlike the one dimensional case, for this two-parameter system our method does not produce estimates that converge to the true parameter values. Figure 5 shows the error surface E_{50} in a small neighborhood of the global minimum. As expected, the sharp gradient discussed in previous sections is in evidence. However, there also seems to be a continuous range of parameter values that generate the same waveform as those located at the global minimum. This alignment is representative of the true behavior of the Hénon system and is not an artifact of the conversion from dynamical system to difference equation. An examination of the gradient of ($y_1[n]$) with respect to the parameters shows that while the ($D_p y_1[n]$) grow quickly for increasing n as expected, they also tend to align themselves along a common axis. The sum of outer products in Eq. 9 is dominated by the matrices formed from these increasingly large gradients that all point in the same direction, and consequently the Hessian appears singular. This behavior is not typical of chaotic dynamical systems in general.

5. CONCLUSION

Using concepts from the discipline of Dynamical Systems we have shown how the sensitive dependence of a chaotic system on its initial conditions can induce an analogous dependence on its parameters. Takens' embedding theorem allowed us to transplant the phase space based notion of Lyapunov exponents which quantify this sensitive depen-

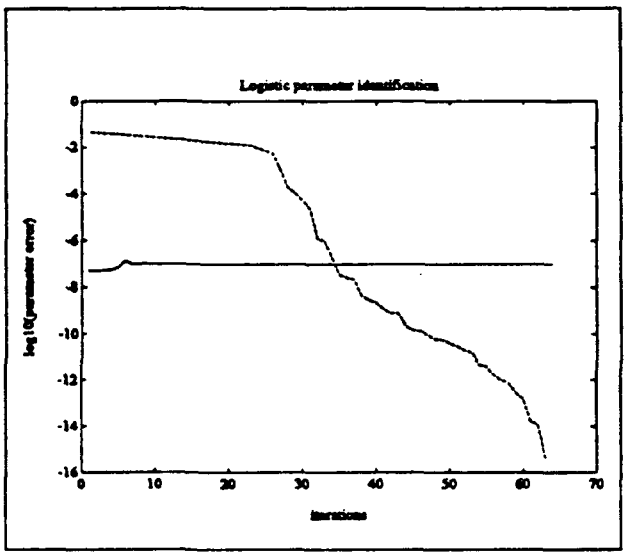

Figure 3: Comparison of waveform error minimization techniques using the chaotic logistic equation. Even good initial parameter estimates get trapped in local minima when trying to minimize the waveform error for the entire target sequence. In contrast, a much poorer initial parameter estimate converges to the true parameter value for a target sequence length of 87 when our modified minimization method is employed.

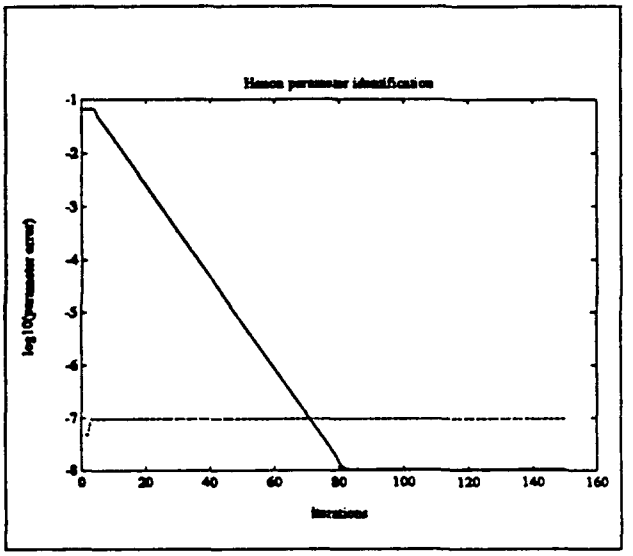

Figure 4: Parameter estimation performance for the Hénon system. Just as in the one-parameter case, the original minimization procedure procedure falls immediately into a local minimum. However, while the modified waveform error method reduces initially much larger parameter deviations better than the original, it does not converge to the true parameters, due to the singular nature of the waveform error's Hessian.

dence into a nonlinear difference equation framework. We explained how such sensitivity could produce conditions ill-suited for a proposed gradient descent minimization of the waveform error, and proposed an improved method to overcome its limitations. The improved method performed significantly better than the original when tested on sequences generated from two chaotic dynamical systems.

6. REFERENCES

[1] J.-C. Roux, R. H. Simoyi, and H. L. Swinney, "Observation of a strange attractor," *Physica D*, vol. 8, pp. 257–266, 1983.

[2] J. Kurths and H. Herzel, "An attractor in a solar time series," *Physica D*, vol. 25, pp. 165–172, 1987.

[3] J. Gleick, *Chaos*. Sphere Books Limited, 1988.

[4] T. F. Quatieri and E. M. Hofstetter, "Short-time signal representation by nonlinear difference equations," in *Proceedings of the 1990 International Conference on Acoustics, Speech, and Signal Processing (Albuquerque,NM)*, 1990.

[5] J.-P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Reviews of Modern Physics*, vol. 57, July 1985.

[6] F. Takens, "Detecting strange attractors in turbulence," in *Lectures Notes in Mathematics, vol. 898*, pp. 366–381, Springer, Berlin, 1981.

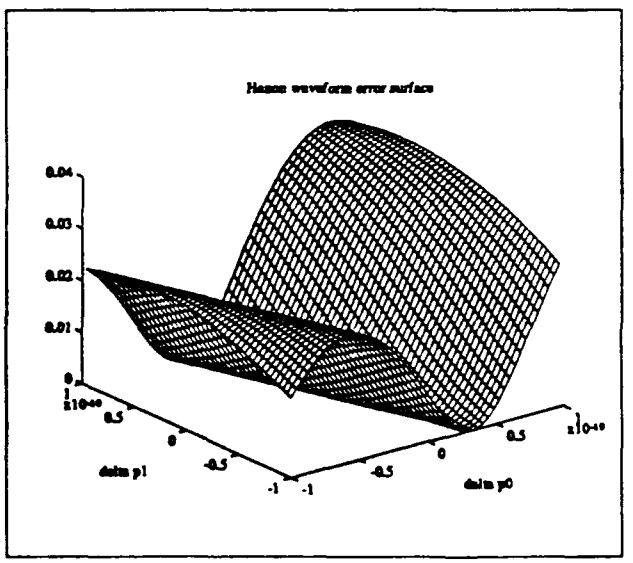

Figure 5: Waveform error surface in the neighborhood of the global minimum. The Hessian appears to be singular.